

Polar Coding for Processes with Memory

Eren Şaşoğlu Intel Corporation

Santa Clara, CA, USA

eren.sasoglu@intel.com

Ido Tal Department of Electrical Engineering

Techion, Haifa 32000, Israel

idotat@ee.technion.ac.il

Abstract

We study polar coding over channels and sources with memory. We show that ψ -mixing processes polarize under the standard transform, and that the rate of polarization to deterministic distributions is roughly $O(2^{-\sqrt{N}})$ as in the memoryless case, where N is the blocklength. This implies that the error probability guarantees of polar channel and source codes extend to a large class of models with memory, including finite-order Markov sources and finite-state channels.

Index Terms

Channels with memory, polar codes, mixing, periodic processes, fast polarization, rate of polarization.

I. INTRODUCTION

Polar codes were invented by Arikan [1] as a low-complexity method to achieve the capacity of symmetric binary-input memoryless channels. The technique that underlies these codes, called *polarization*, is quite versatile, and has since been applied to numerous classical memoryless problems in information theory.

Many practical sources and channels are not well-described by memoryless models. In wireless communication, for example, memory in the form of intersymbol interference is quite prominent due to multipath propagation, as are slow variations in channel conditions due to mobility. In practice, this type of memory is most commonly handled by eliminating it, by augmenting the transmitter/receiver appropriately to create an overall memoryless channel. Memoryless coding techniques are then used for communication. Channel equalization and OFDM techniques are perhaps the most notable examples of this approach.

In contrast, we are interested here in whether polar coding can be used *directly* on channels and sources with memory. In addition to being of theoretical interest, such results may help simplify the design of communication or compression systems.

Little is known about the theory of polarization for settings with memory. In particular, it was shown in [2] that the successive cancellation decoding complexity of polar codes scales with the number of states of the underlying process, and thus is practical if the amount of memory in the system is modest. It was shown in [3, Chapter 5] that Arikan's standard transform indeed polarizes strongly mixing processes with finite memory. Whether polarization takes place sufficiently fast to yield a coding theorem has been left open, however, and that is the problem we address here.

We first give a proof of polarization that is both simpler than the one given in [3], and holds for the more general class of ψ -mixing processes. We then show that the asymptotic rate of polarization to deterministic distributions is as in the memoryless case. This lets us conclude that the usual asymptotic error probability guarantees of polar channel and source codes carry over to processes with memory, including well-behaved Markov sources as well as the practical channel models mentioned above.

II. SETTING

Let (X_i, Y_i, S_i) , $i \in \mathbb{Z}$, be a stationary process, where Y_i and S_i take values in finite alphabets \mathcal{Y} and \mathcal{S} . We assume $X_i \in \{0, 1\}$ in order to keep the notation simple, but the results here can be generalized to arbitrary finite alphabets using standard techniques. See, for example, [3, Chapter 3].

We think of X_i as a sequence to be estimated, and Y_i as a sequence of observations related to X_i . In particular, X_i may be the input sequence to a communication channel, with the corresponding channel output Y_i . Alternatively, X_i may be the output of a data source to be compressed, and Y_i may be the side information available to the decompressor. A (possibly unknown) state sequence S_i may underlie the channel or the source. Frequently, it is assumed that the pair (X_i, Y_i) is independent of the history $(X_1^{i-1}, Y_1^{i-1}, S_1^{i-1})$ of the process conditioned on the present state S_i .

We assume throughout that the process (X_i, Y_i, S_i) is ψ -mixing. We follow¹ [4, Page 169] and say that a process T_i is ψ -mixing if there exists a sequence $\psi_k \rightarrow 1$ as $k \rightarrow \infty$ such that

$$\Pr(A \cap B) \leq \psi_k \Pr(A) \Pr(B) \quad (1)$$

This work was done when Eren Şaşoğlu was at the Technion in June–July 2015. It was presented in part at ISIT 2016.

¹To the best of our understanding, the first displayed equation on page 169 of [4] should be “ $\sum_v \mu(uvv) \leq \dots$ ”.

for all $A \in \sigma(T_{-\infty}^0)$ and $B \in \sigma(T_{k+1}^\infty)$, where $\sigma(\cdot)$ denotes the sigma-field generated by its argument. Note that the dependence of ψ_k on events A and B is only through the distance k between them. Therefore, in a ψ -mixing process, any two events that are sufficiently separated in ‘time’ are almost independent.

Many source and channel models of practical importance are captured by ψ -mixing. In particular,

- (i) an independent and identically distributed (i.i.d.) data source X_i is ψ -mixing.
- (ii) A finite-order, stationary, irreducible, aperiodic Markov source X_i is ψ -mixing.
- (iii) Suppose X_i is a stationary data source with state S_i , where the next source symbol and state depend only on their current values. That is,

$$p(s_{i+1}, x_i | s_{-\infty}^i, x_{-\infty}^{i-1}) = p(s_{i+1}, x_i | s_i, x_{i-1}) .$$

The process (S_i, X_i) is Markov. If it is also irreducible and aperiodic, then by (ii) it is ψ -mixing, and therefore so is X_i . This model includes data sources generated by a hidden Markov state sequence, described by the conditional distributions

$$p(s_i, x_i | s_{-\infty}^{i-1}, x_{-\infty}^{i-1}) = p(s_i | s_{i-1})p(x_i | s_i) .$$

- (iv) If X_i is an i.i.d. input sequence to a discrete memoryless channel and Y_i is the corresponding channel output sequence, then (X_i, Y_i) is i.i.d. and therefore ψ -mixing by (i).
- (v) Suppose W is a finite-state channel with input sequence X_i , output sequence Y_i , and state sequence S_i , all taking values in finite but otherwise arbitrary sets. The current output and the next state of the channel depend only on the current state and input [5]:

$$p(s_i, y_i | x_{-\infty}^{i-1}, s_{-\infty}^{i-1}, y_{-\infty}^{i-1}) = W(s_i, y_i | x_{i-1}, s_{i-1}) .$$

If the input X_i is a Markov process, then so is (X_i, Y_i, S_i) , and thus the latter is also ψ -mixing.

The parameter ψ_0 plays an important role in this paper, and can be computed easily for all of the cases above [4, Page 169].

We are interested in the effects of Arkan’s standard polar transform on ψ -mixing processes. For this purpose, we let $U_1^N = X_1^N \mathbf{B}_N \mathbf{G}_N$, where the matrix multiplications are over the binary field, $N = 2^n$ for positive integers n , \mathbf{G}_N is the n th Kronecker power of $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, and \mathbf{B}_N is the $N \times N$ bit-reversal matrix. The conditional entropy rate of X_i is defined as

$$\mathcal{H}_{X|Y} = \lim_{N \rightarrow \infty} \frac{1}{N} H(X_1^N | Y_1^N) = \lim_{N \rightarrow \infty} \frac{1}{N} H(X_1^N, Y_1^N) - \lim_{N \rightarrow \infty} \frac{1}{N} H(Y_1^N) .$$

The limits on the right-hand-side exist due to stationarity [6, Theorem 4.2.1]. Also useful for the analysis is the parameter

$$Z(A|B) = 2 \sum_{b \in \mathcal{B}} \sqrt{p_{AB}(0, b) p_{AB}(1, b)}$$

for random variables $A \in \{0, 1\}$ and B . Sometimes called the Bhattacharyya parameter, $Z(A|B)$ upper bounds the error probability of optimally guessing A by observing B . See, for example [3, Proposition 2.2].

III. MAIN RESULTS

Theorem 1 (Polarization). *If $\psi_0 < \infty$, then for all $\epsilon > 0$*

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} |\{i : H(U_i | U_1^{i-1} Y_1^N) > 1 - \epsilon\}| &= \mathcal{H}_{X|Y}, \\ \lim_{N \rightarrow \infty} \frac{1}{N} |\{i : H(U_i | U_1^{i-1} Y_1^N) < \epsilon\}| &= 1 - \mathcal{H}_{X|Y}. \end{aligned}$$

Theorem 2 (Fast polarization). *If $\psi_0 < \infty$, then for all $\beta < 1/2$*

$$\lim_{N \rightarrow \infty} \frac{1}{N} |\{i : Z(U_i | U_1^{i-1} Y_1^N) < 2^{-N^\beta}\}| = 1 - \mathcal{H}_{X|Y} .$$

Theorem 3 (Periodic processes may not polarize). *The stationary periodic Markov process described in Figure 1 does not polarize. Indeed, for all $\frac{5N}{8} < i \leq \frac{6N}{8}$,*

$$\left| H(U_i | U_1^{i-1}) - \frac{1}{2} \right| \leq \epsilon_N, \quad \lim_{N \rightarrow \infty} \epsilon_N = 0 . \quad (2)$$

We will prove these claims in the following sections. Throughout, we will use the shorthand

$$\begin{aligned} H^{\mathbf{b}} &= H(U_i | Y_1^N U_1^{i-1}), \\ Z^{\mathbf{b}} &= Z(U_i | Y_1^N U_1^{i-1}), \end{aligned}$$

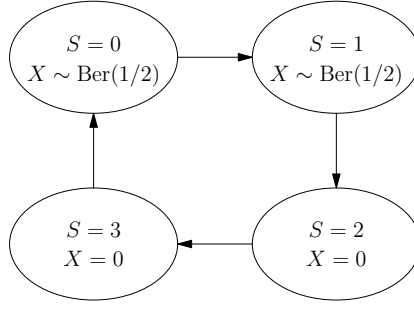


Fig. 1. A periodic data source that does not polarize. The source output is Bernoulli 1/2 for two consecutive states and zero for next two consecutive states. There is no side information, i.e., Y_i is constant.

where $\mathbf{b} \in \{0, 1\}^n$ is the n -bit binary expansion of $i - 1 \in \{0, \dots, N - 1\}$. We will omit the ranges of indices when they are clear from context. The following are immediate from the definition of $\mathbf{B}_N \mathbf{G}_N$:

$$\begin{aligned} H^{\mathbf{b}0} &= H(U_{2i-1} | Y_1^{2N} U_1^{2i-2}) \\ H^{\mathbf{b}1} &= H(U_{2i} | Y_1^{2N} U_1^{2i-1}) \end{aligned}$$

for all $\mathbf{b} \in \{0, 1\}^n$. These identities also hold when the H 's are replaced by Z 's.

Further, if we let B_1, B_2, \dots be a sequence of i.i.d. $\text{Ber}(1/2)$ random variables, then it is easy to see that the random variables $H_n = H^{B_1 \dots B_n}$ and $Z_n = Z^{B_1 \dots B_n}$ are uniformly distributed over the sets of $H^{\mathbf{b}}$'s and $Z^{\mathbf{b}}$'s, respectively. Theorems 1 and 2 are then equivalent to

Theorem 4. *If $\psi_0 < \infty$, then for all $\epsilon > 0$*

$$\begin{aligned} \lim_{n \rightarrow \infty} P(H_n > 1 - \epsilon) &= \mathcal{H}_{X|Y}, \\ \lim_{n \rightarrow \infty} P(H_n < \epsilon) &= 1 - \mathcal{H}_{X|Y}. \end{aligned}$$

Theorem 5. *If $\psi_0 < \infty$, then for all $\beta < 1/2$*

$$\lim_{n \rightarrow \infty} P(Z_n < 2^{-N^\beta}) = 1 - \mathcal{H}_{X|Y}.$$

As is usual in proofs of polarization, we will analyze how the entropies and Bhattacharyya parameters evolve in a single recursion of the polarization transform, that is, when two smaller polarization blocks are combined to form a larger block. Due to the dependence between the combined blocks, we will need to keep track of more random variables than is required in the analysis of the memoryless case. The following shorthand will then be useful:

$$\begin{aligned} U_1^N &= X_1^N \mathbf{B}_N \mathbf{G}_N \\ V_1^N &= X_{N+1}^{2N} \mathbf{B}_N \mathbf{G}_N \\ Q_i &= Y_1^N U_1^{i-1} \\ R_i &= Y_{N+1}^{2N} V_1^{i-1} \end{aligned} \tag{3}$$

IV. PROOF OF THEOREM 1

For the proof, we take the standard approach of showing that H_n converges almost surely to a $\{0, 1\}$ -valued random variable. Recall that we have defined H_n through

$$H_n = H(U_i | Y_1^N U_1^{i-1}) = H(U_i | Q_i) \quad \text{whenever } (B_1 \dots B_n)_2 = i - 1.$$

For a given realization of H_n , observe that we have

$$H_{n+1} = \begin{cases} H(U_i + V_i | Q_i, R_i) & \text{if } B_{n+1} = 0 \\ H(V_i | Q_i, R_i, U_i + V_i) & \text{if } B_{n+1} = 1 \end{cases}.$$

Along with these identities, the relation

$$H(U_i + V_i | Q_i, R_i) + H(V_i | Q_i, R_i, U_i + V_i) = H(U_i, V_i | Q_i, R_i) \leq 2H(U_i | Q_i)$$

implies that $E[H_{n+1}|H_1, \dots, H_n] \leq H_n$. Since the entropy is bounded as $H_n \in [0, 1]$, it follows that H_1, H_2, \dots is a bounded supermartingale and thus converges almost surely to a $[0, 1]$ -valued random variable H_∞ . It therefore remains to show that $H_\infty \in \{0, 1\}$ almost surely. For this purpose, we will show that for all $\xi > 0$ there exists $\gamma(\xi) > 0$ such that

$$H(U_i|Q_i) \in (2\xi, 1 - 2\xi) \text{ implies } H(U_i + V_i|Q_i, R_i) - H(U_i|Q_i) > \gamma(\xi) \text{ for almost all } i, \quad (4)$$

that is, for a fraction of $i \in \{1, \dots, N\}$ approaching 1 as $N \rightarrow \infty$. The theorem will follow from this claim, since (4) is equivalent to saying that if H_n is bounded away from 0 and 1, then $H_{n+1} - H_n$ is almost surely bounded away from 0. Therefore whenever H_n converges, it can do so only to 0 or 1.

In the rest of this section, we show (4). We know from [3, Chapter 3] that the claim would hold for all N and i if (X_1^N, Y_1^N) and $(X_{N+1}^{2N}, Y_{N+1}^{2N})$ were independent. Our purpose here is to show that in the present setting there is sufficient independence between various random variables in neighboring blocks to imply (4). (This is essentially the approach taken in [3, Chapter 5], although the proof here is simpler and more general.) In particular, we will need the following independence results.

Lemma 6. *If $\psi_0 < \infty$, then for any $\epsilon > 0$, the fraction of indices i for which*

$$I(U_i; R_i|Q_i) < \epsilon \quad (5)$$

$$I(V_i; Q_i|R_i) < \epsilon \quad (6)$$

$$I(U_i; V_i|Q_i, R_i) < \epsilon \quad (7)$$

approaches 1 as $N \rightarrow \infty$.

Proof: We only prove the first and the third inequalities, the second follows by symmetry. We have

$$\begin{aligned} \log(\psi_0) &\geq E \left[\log \frac{p_{X_1^{2N} Y_1^{2N}}}{p_{X_1^N Y_1^N} \cdot p_{X_{N+1}^{2N} Y_{N+1}^{2N}}} \right] \\ &= I(X_1^N Y_1^N; X_{N+1}^{2N} Y_{N+1}^{2N}) \\ &= I(U_1^N Y_1^N; V_1^N Y_{N+1}^{2N}) \\ &= I(Y_1^N; V_1^N Y_{N+1}^{2N}) + I(U_1^N; V_1^N Y_{N+1}^{2N} | Y_1^N) \\ &\geq I(U_1^N; V_1^N Y_{N+1}^{2N} | Y_1^N) \\ &= \sum_{i=1}^N I(U_i; V_1^N Y_{N+1}^{2N} | Y_1^N U_1^{i-1}) \\ &= \sum_{i=1}^N I(U_i; V_i R_i V_{i+1}^N | Q_i), \end{aligned}$$

The first inequality above follows from the definition of ψ_0 . Since all terms inside the last sum are non-negative, it follows that at most $\sqrt{\log(\psi_0)N}$ (a vanishing fraction) of them are at most $\sqrt{\log(\psi_0)/N}$. Thus, to conclude the proof, it suffices to show that the i th term is greater than both $I(U_i; R_i|Q_i)$ and $I(U_i; V_i|Q_i, R_i)$. Indeed,

$$I(U_i; V_i R_i V_{i+1}^N | Q_i) = I(U_i; R_i|Q_i) + I(U_i; V_i|Q_i R_i) + I(U_i; V_{i+1}^N | Q_i V_i R_i),$$

and all the terms are non-negative. ■

Lemma 6 shows “almost-independence” between various pairs of random variables from neighboring blocks. However, since Y_N and Y_{N+1} are generally dependent, we cannot hope to make a similar claim of independence between Q_i and R_i . This presents an apparent problem: suppose Q_i is such that $H(U_i|Q_i = q_i)$ is either 0 or 1 depending on q_i , although the average entropy $H(U_i|Q_i)$ is bounded away from both 0 and 1. By stationarity, it must also be that $H(V_i|R_i = r_i)$ is also either 0 or 1. Now, since Q_i and R_i are not independent, it is conceivable that they collude: with probability 1, we either have $H(U_i|Q_i = q_i) = H(V_i|R_i = r_i) = 0$ or $H(U_i|Q_i = q_i) = H(V_i|R_i = r_i) = 1$. If furthermore all the mutual informations in Lemma 6 are equal to 0, then no polarization occurs: both $H(U_i + V_i|Q_i R_i)$ and $H(V_i|U_i + V_i, Q_i R_i)$ equal $H(U_i|Q_i)$. We will use the following lemma later to show that such harmful collusions are, in fact, impossible.

Lemma 7. *Let (X_i, Y_i) be stationary and ψ -mixing. For all $\xi > 0$, there exists N_0 and $\delta(\xi) > 0$ such that for all $N > N_0$ and all $\{0, 1\}$ -valued random variables $A = f(X_1^N, Y_1^N)$ and $B = f(X_{N+1}^{2N}, Y_{N+1}^{2N})$*

$$p_A(0) \in (\xi, 1 - \xi) \text{ implies } p_{AB}(0, 1) > \delta(\xi).$$

Proof: Let us start by explaining informally why the claim is true. Define $C = f(X_{2N+1}^{3N}, Y_{2N+1}^{3N})$, and suppose to the contrary that B equals A with very high probability. Hence, by stationarity, C equals B with very high probability. We conclude that A equals C with probability very close to 1, a contradiction to the mixing property.

Let us now give a formal proof. We have

$$\begin{aligned}
2p_{AB}(0, 1) &= p_{AB}(0, 1) + p_{BC}(0, 1) \\
&\geq p_{ABC}(0, 1, 1) + p_{ABC}(0, 0, 1) \\
&= p_{AC}(0, 1) \\
&= p_A(0) - p_{AC}(0, 0) \\
&\geq p_A(0)(1 - \psi_N p_C(0)) \\
&= p_A(0)(1 - \psi_N p_A(0))
\end{aligned}$$

where the first and last equalities are due to stationarity. Since $p_A(0) \in (\xi, 1 - \xi)$ and $\psi_N \rightarrow 1$, it follows that there exists an N_0 such that the last term is away from 0 for all $N > N_0$. Further, since ψ_N is independent of A , so is N_0 . This yields the claim. \blacksquare

We are almost ready to complete the proof by showing (4). The final simplification we will need is to approximate the distribution of (U_i, V_i, Q_i, R_i) by another distribution that is easier to work with. In particular, let $(\tilde{U}_i, \tilde{V}_i)$ be random variables whose joint distribution with (Q_i, R_i) is of the form

$$p_{\tilde{U}_i \tilde{V}_i Q_i R_i}(u_i, v_i, q_i, r_i) = p_{U_i | Q_i}(u_i | q_i) p_{V_i | R_i}(v_i | r_i) p_{Q_i R_i}(q_i, r_i). \quad (8)$$

Consider (5)–(7), and note that if we replace U_i and V_i by \tilde{U}_i and \tilde{V}_i , respectively, then all the mutual informations become zero; see (20)–(22) in the appendix for a proof of this fact. The closeness between the two distributions is harnessed in the following corollary to Lemma 6 (proved in the appendix):

Corollary 8. *If $\psi_0 < \infty$, then for any $\epsilon > 0$, the fraction of indices i for which*

$$|H(\tilde{U}_i + \tilde{V}_i | Q_i R_i) - H(U_i + V_i | Q_i R_i)| < \epsilon$$

approaches 1 as $N \rightarrow \infty$.

By (8), we have

$$H(\tilde{U}_i | Q_i R_i) = H(\tilde{U}_i | Q_i) = H(U_i | Q_i).$$

Observe that we need to show (4) for arbitrarily small *but fixed* ξ . Therefore, it suffices to show that

$$H(\tilde{U}_i | Q_i R_i) \in (2\xi, 1 - 2\xi) \text{ implies } H(\tilde{U}_i + \tilde{V}_i | Q_i R_i) - H(\tilde{U}_i | Q_i R_i) > 2\gamma(\xi) \quad (9)$$

for almost all i in order to complete the proof. To do so, we will use the following fact, whose proof is given in the appendix.

Lemma 9. *Let A and B be independent binary random variables. For every $\xi > 0$, there exists $\Delta(\xi) > 0$ such that*

$$\max\{H(A), H(B)\} > \xi \quad \text{and} \quad \min\{H(A), H(B)\} < 1 - \xi$$

imply

$$H(A + B) > \frac{H(A) + H(B)}{2} + \Delta(\xi).$$

For a given i , define the random variables $H_{Q_i}(\tilde{U}_i)$, $H_{R_i}(\tilde{V}_i)$, and $H_{Q_i R_i}(\tilde{U}_i + \tilde{V}_i)$ that take the values

$$\begin{aligned}
H_{Q_i}(\tilde{U}_i) &= H(\tilde{U}_i | Q_i = q_i) \\
H_{R_i}(\tilde{V}_i) &= H(\tilde{V}_i | R_i = r_i) \\
H_{Q_i R_i}(\tilde{U}_i + \tilde{V}_i) &= H(\tilde{U}_i + \tilde{V}_i | (Q_i, R_i) = (q_i, r_i))
\end{aligned}$$

whenever

$$(Q_i, R_i) = (q_i, r_i).$$

Note that (9) is equivalent to:

$$E[H_{Q_i}(\tilde{U}_i)] \in (2\xi, 1 - 2\xi) \text{ implies } E[H_{Q_i R_i}(\tilde{U}_i + \tilde{V}_i) - H_{Q_i}(\tilde{U}_i)] \geq 2\gamma(\xi).$$

We take $2\gamma(\xi) = \delta(\xi)\Delta(\xi)$, where $\delta(\xi)$ and $\Delta(\xi)$ are as in Lemmas 7 and 9, respectively. We will be done if we can show that $E[H_{Q_i}(\tilde{U}_i)] \in (2\xi, 1 - 2\xi)$ implies

$$P\left(\max\{H_{Q_i}(\tilde{U}_i), H_{R_i}(\tilde{V}_i)\} > \xi \quad \text{and} \quad \min\{H_{Q_i}(\tilde{U}_i), H_{R_i}(\tilde{V}_i)\} < 1 - \xi\right) > \delta(\xi). \quad (10)$$

Indeed, Lemma 9 and stationarity then imply that

$$E[H_{Q_i R_i}(\tilde{U}_i + \tilde{V}_i) - H_{Q_i}(\tilde{U}_i)] = E\left[H_{Q_i R_i}(\tilde{U}_i + \tilde{V}_i) - \frac{H_{Q_i}(\tilde{U}_i) + H_{R_i}(\tilde{V}_i)}{2}\right] \geq \delta(\xi)\Delta(\xi).$$

Let us assume without loss of generality that $\delta(\xi) < \xi$. Thus, if $P(H_{Q_i}(\tilde{U}_i) \in (\xi, 1 - \xi)) \geq \xi$, then (10) is immediate. Let us suppose then that

$$P(H_{Q_i}(\tilde{U}_i) \in (\xi, 1 - \xi)) < \xi.$$

Since $H_{Q_i}(\tilde{U}_i) \in [0, 1]$ and $E[H_{Q_i}(\tilde{U}_i)] \in (2\xi, 1 - 2\xi)$, it follows by Markov's inequality that

$$P(H_{Q_i}(\tilde{U}_i) > 1 - \xi) \in \left(\frac{\xi}{1 - \xi}, \frac{1 - 2\xi}{1 - \xi} \right) \subseteq (\xi, 1 - \xi).$$

Further, there exists a function f such that $\mathbb{1}_{[H_{Q_i}(\tilde{U}_i) > 1 - \xi]} = f(X_1^N, Y_1^N)$ and $\mathbb{1}_{[H_{R_i}(\tilde{V}_i) > 1 - \xi]} = f(X_{N+1}^{2N}, Y_{N+1}^{2N})$. It therefore follows from Lemma 7 that

$$P(H_{Q_i}(\tilde{U}_i) > 1 - \xi, H_{R_i}(\tilde{V}_i) \leq 1 - \xi) > \delta(\xi),$$

implying (10). This completes the proof.

V. PROOF OF THEOREM 2

Like most proofs of the speed of polarization, our proof of Theorem 2 relies on the following result by Arıkan and Telatar [7], although we need the more general form of the result given in [3, Lemma 2.3].

Lemma 10 ([7],[3]). *If Z_n converges almost surely to a random variable Z_∞ and if there exists $K < \infty$ such that*

$$Z_n \leq K Z_{n-1} \quad \text{if } B_n = 0 \quad (11)$$

$$Z_n \leq K Z_{n-1}^2 \quad \text{if } B_n = 1 \quad (12)$$

then

$$\lim_{n \rightarrow \infty} P(Z_n < 2^{-2^{n\beta}}) = P(Z_\infty = 0)$$

for all $\beta < 1/2$.

Recall from the proof of Theorem 1 that H_n almost surely converges to a $\{0, 1\}$ -valued random variable. It then follows from the relations [8]

$$Z(A|B)^2 \leq H(A|B)$$

$$H(A|B) \leq \log(1 + Z(A|B))$$

that Z_n also converges almost surely, and in particular $Z_n \rightarrow 0$ whenever $H_n \rightarrow 0$, and $Z_n \rightarrow 1$ whenever $H_n \rightarrow 1$. It then suffices to show that Z_n satisfies inequalities (11) and (12).

We claim that this is indeed the case with $K = 2\psi_0$. To see this, let $\hat{X}_1^{2N}, \hat{Y}_1^{2N}$ be distributed as $P_{X_1^N Y_1^N} \cdot P_{X_{N+1}^{2N} Y_{N+1}^{2N}}$, and define the corresponding variables $\hat{U}_i, \hat{V}_i, \hat{Q}_i, \hat{R}_i$ as in (3). We know from [1] that

$$Z(\hat{U}_i + \hat{V}_i | \hat{Q}_i, \hat{R}_i) \leq 2Z(\hat{U}_i | \hat{Q}_i) \quad (13)$$

$$Z(\hat{V}_i | \hat{Q}_i, \hat{R}_i, \hat{U}_i + \hat{V}_i) \leq Z(\hat{U}_i | \hat{Q}_i)^2 \quad (14)$$

Now let (A, B) and (\hat{A}, \hat{B}) be random variables that can be written as

$$(A, B) = f(X_1^{2N}, Y_1^{2N})$$

$$(\hat{A}, \hat{B}) = f(\hat{X}_1^{2N}, \hat{Y}_1^{2N})$$

for some function f . Observe that the assumption (1) implies $P_{AB} \leq \psi_0 \cdot P_{\hat{A}\hat{B}}$. Therefore, for binary A we have

$$\begin{aligned} Z(A|B) &= 2 \sum_b \sqrt{p_{AB}(0, b) p_{AB}(1, b)} \\ &\leq 2\psi_0 \sum_b \sqrt{p_{\hat{A}\hat{B}}(0, b) p_{\hat{A}\hat{B}}(1, b)} \\ &= \psi_0 \cdot Z(\hat{A}|\hat{B}) \end{aligned} \quad (15)$$

Defining $A = U_i + V_i$ and $B_i = (Q_i, R_i)$ and combining (15) with (13) implies (11) with $K = 2\psi_0$. Similarly, defining $A = V_i$ and $B_i = (Q_i, R_i, U_i + V_i)$ and combining (15) with (14) implies (12) with $K = \psi_0$. This proves Theorem 2 since $\psi_0 < \infty$ by assumption.

VI. PROOF OF THEOREM 3

Recall that the process we are considering is described in Figure 1. Let us start by defining the process exactly. The state of the process at time $t = 1, 2, \dots$ is denoted S_t . Each such state has 4 possible values, $\{0, 1, 2, 3\}$. The initial state S_1 is picked uniformly at random. The value of S_1 determines the value of all S_t , specifically, $S_t = S_1 + (t - 1) \bmod 4$. If $S_t \in \{0, 1\}$, then X_t , the output of the process at time t , is picked uniformly at random from $\{0, 1\}$. If $S_t \in \{2, 3\}$, then X_t equals 0. Recall that for a given N , we have $U_1^N = X_1^N \mathbf{B}_N \mathbf{G}_N$.

The proof of Theorem 3 is divided into two parts. In the first part, we consider $H(U_i | U_1^{i-1}, S_1 = s_1)$. Namely, we consider a setting related to yet distinct from that of Theorem 3: we assume that the initial state S_1 is known to equal the fixed value s_1 . As we will see, the case $N = 8$ is of particular importance. We refer the reader to Table II, which highlights key features of the distribution of U_1^6 when $N = 8$, for the 4 possible values of s_1 . The entry " $U_6 \perp U_1^5$ " denotes that U_6 is independent of U_1^5 . The correctness of the Table II is easy to validate by using Table I.

Lemma 11. *Consider the stationary Markov process described in Figure 1. Then, for $N \geq 8$, the following holds.*

$$\text{For all } \frac{5N}{8} < i \leq \frac{6N}{8} \text{ we have that } H(U_i | U_1^{i-1}, S_1 = s_1) = \begin{cases} 0 & \text{if } s_1 \in \{1, 3\} \\ 1 & \text{if } s_1 \in \{0, 2\} \end{cases}.$$

Proof: The correctness of the lemma is straightforward to validate for $N = 8$. Indeed, for $N = 8$ we must only consider $i = 6$, and the result follows from the last column of Table II. Namely, for $s_1 \in \{1, 3\}$ we have that U_6 is a function of U_1^5 ; for $s_1 \in \{0, 2\}$ we have that U_6 is independent of U_1^5 and is distributed $\text{Ber}(1/2)$.

The general result is proved by induction on N . We have proved the basis $N = 8$ above. In order to prove the step, let us first tailor the notation (3) to our needs:

$$\begin{aligned} U_1^N &= X_1^N \mathbf{B}_N \mathbf{G}_N \\ V_1^N &= X_{N+1}^{2N} \mathbf{B}_N \mathbf{G}_N \\ Q_i &= U_1^{i-1} \\ R_i &= V_1^{i-1} \end{aligned} \tag{16}$$

Proving the step is equivalent to proving that for all $\frac{5N}{8} < i \leq \frac{6N}{8}$,

$$H(U_i + V_i | Q_i R_i, S_1 = s_1) = H(V_i | U_i + V_i, Q_i R_i, S_1 = s_1) = H(U_i | Q_i, S_1 = s_1). \tag{17}$$

Recall that N is a power of 2 and $N \geq 8$. Thus, N is a multiple of 4. Since the period of the process is 4, we have that $S_1 = s_1$ iff $S_{N+1} = s_1$. Moreover, it is easily seen that (U_i, Q_i) is independent of (V_i, R_i) . Hence,

$$H(U_i | Q_i, S_1 = s_1) = H(V_i | R_i, S_{N+1} = s_1) = H(V_i | R_i, S_1 = s_1). \tag{18}$$

It is now simple to prove (17) from (18) for the two cases of interest. Indeed, if $H(U_i | Q_i, S_1 = s_1) = 0$ then U_i and V_i are deterministic function of Q_i and R_i , respectively, given that $S_1 = s_1$. Hence, the two equalities in (17) follow easily. If $H(U_i | Q_i, S_1 = s_1) = 1$ then, given that $S_1 = s_1$, we have that U_i and V_i are independent, each one distributed $\text{Ber}(1/2)$. Moreover, $(U_i V_i)$ and $(Q_i R_i)$ are independent, given that $S_1 = s_1$. Again, (17) follows easily. ■

An immediate corollary of Lemma 11 is that $H(U_i | U_1^{i-1}, S_1) = 1/2$, for $\frac{5N}{8} < i \leq \frac{6N}{8}$. To see this, note that all 4 states are equally likely as initial states. What remains is to prove that S_1 is essentially known from U_1^{i-1} .

Lemma 12. *Consider the stationary Markov process depicted in Figure 1. Then, there exists an ϵ_N such that*

$$\text{for all } \frac{5N}{8} < i \leq \frac{6N}{8} \text{ we have that } H(S_1 | U_1^{i-1}) \leq \epsilon_N, \quad \text{and} \quad \lim_{N \rightarrow \infty} \epsilon_N = 0. \tag{19}$$

Proof: We start by giving an informal explanation as to why the claim holds. Consider the first two columns of Table II, and suppose we had many i.i.d. realizations of U_1^5 , all with the same initial state s_1 . Hence, the first column would allow us to distinguish — with very high probability — between $s_1 = 0$, $s_1 = 2$, and $s_1 \in \{1, 3\}$:

- If $s_1 = 0$ then all the realizations of U_4 would equal 0.
- If $s_1 = 2$, all realizations would satisfy $U_2 = U_4$. In roughly half the realizations we would have $U_4 = 1$, since $U_4 \sim \text{Ber}(1/2)$. Each such realization would rule out the previous case.
- If $s_1 \in \{1, 3\}$ then in roughly a quarter of the realizations we would have $U_4 = 1$ and $U_2 = 0$, since U_2 and U_4 are i.i.d. and $\text{Ber}(1/2)$. Such an outcome would distinguish this case from the two previous ones.

To distinguish between $s_1 = 1$ and $s_1 = 3$, we utilize the second column of Table II. Specifically, in both cases, $U_1 \sim \text{Ber}(1/2)$. Thus, in roughly half of the realizations, $U_1 = 1$, and for each such realization we can distinguish between $s_1 = 1$ in which $U_5 = U_3$ and $s_1 = 3$ in which $U_5 \neq U_3$.

Lastly, we claim that such independent realization of U_1^5 can indeed be attained. Specifically, for $N \geq 8$ and $\frac{5N}{8} < i \leq \frac{6N}{8}$, the vector U_1^{i-1} can be used to deduce the first 5 entries of each vector in the set $\{X_{1+8(j-1)}^{1+8j} \mathbf{B}_8 \mathbf{G}_8 : 1 \leq j \leq N/8\}$. Note

$U_1 =$	X_1	$+$	X_2	$+$	X_3	$+$	X_4	$+$	X_5	$+$	X_6	$+$	X_7	$+$	X_8
$U_2 =$									X_5	$+$	X_6	$+$	X_7	$+$	X_8
$U_3 =$				X_3	$+$	X_4							X_7	$+$	X_8
$U_4 =$													X_7	$+$	X_8
$U_5 =$			X_2			$+$	X_4			$+$	X_6			$+$	X_8
$U_6 =$											X_6			$+$	X_8
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8							
$S_1 = 0$	B	B	0	0	B	B	0	0							
$S_1 = 1$	B	0	0	B	B	0	0	B							
$S_1 = 2$	0	0	B	B	0	0	B	B							
$S_1 = 3$	0	B	B	0	0	B	B	0							

TABLE I

PROPERTIES OF U_1^6 AND X_1^8 FOR $N = 8$. UPPER HALF: U_1^6 AS A FUNCTION OF X_1^8 . LOWER HALF: DISTRIBUTION OF X_1^8 AS A FUNCTION OF THE INITIAL STATE S_1 . IN THE LOWER HALF, “ B ” IS SHORT FOR $\text{Ber}(1/2)$ AND “ 0 ” DESIGNATES A VALUE OF ZERO WITH PROBABILITY ONE.

	(U_2, U_4)	(U_1, U_3, U_5)	U_6 vs. U_1^5
$S_1 = 0$	$U_4 = 0$		$U_6 \perp U_1^5$
$S_1 = 1$	i.i.d.	$U_5 = U_3$	$U_6 = U_4$
$S_1 = 2$	$U_4 = U_2$		$U_6 \perp U_1^5$
$S_1 = 3$	i.i.d.	$U_5 = U_3 + U_1$	$U_6 = U_4 + U_2$

TABLE II

DISTRIBUTION PROPERTIES OF U_1^6 FOR $N = 8$ AND THE FOUR POSSIBLE INITIAL STATES.

that since the period of the process is 4, the state at time $1 + 8(j - 1)$ is equal to s_1 , for all values of j . Also, given S_1 , all the vectors in the above set are independent.

Let us move on to the formal proof. The statistical properties of U_1^5 detailed above are easy to validate using Table I. Suppose we have $N/8$ realizations of U_1^5 , which are i.i.d. given S_1 . The above description suggests an algorithm for guessing the value of S_1 :

- If all the realizations of U_4 equal 0, set $\hat{S}_1 = 0$.
- Otherwise, if all realizations satisfy $U_2 = U_4$, set $\hat{S}_1 = 2$.
- Otherwise, if all realizations satisfy $U_5 = U_3$, set $\hat{S}_1 = 1$.
- Otherwise, set $\hat{S}_1 = 3$.

A straightforward calculation shows that the probability of misdecoding S_1 goes down to 0 exponentially in N . By Fano’s inequality [6, Theorem 2.10.1], we have that

$$H(S|U_0^{i-1}) \leq h(p_e) + p_e \log_2 3 ,$$

where p_e is the probability of misdecoding. Since p_e tends to 0, the RHS of the above tends to 0 as well.

Recall the set $\{X_{1+8(j-1)}^{1+8j} B_8 G_8 : 1 \leq j \leq N/8\}$, and denote by A the vectors obtained by taking the prefix of length 5 of each vector in the set. Obviously, the vectors in A are i.i.d. given S_1 , and have the same distribution as the U_1^5 discussed above. All that remains to prove is that we can deduce A from U_1^{i-1} , when $\frac{5N}{8} < i \leq \frac{6N}{8}$. We prove this by induction on N . The case $N = 8$ is immediate. For the step, let the set B be defined similarly to A , but with j ranging as $N/8 + 1 \leq j \leq N/8 + N/8$. The induction step assumes that A can be deduced from U_1^{i-1} . Hence, B can be deduced from V_1^{i-1} , where we recall the shorthand (16). Recalling the definition of the polar transform, we must prove that both A and B can be deduced from either $(U_1^{i-1} + V_1^{i-1}, V_1^{i-1})$ or $(U_1^{i-1} + V_1^{i-1}, V_1^{i-1}, U_i + V_i)$. Obviously, this is true. ■

The proof of Theorem 3 is now a simple consequence of the above.

Proof of Theorem 3: By the chain rule applied in two ways to $H(U_i S_1 | U^{i-1})$ we deduce that

$$H(U_i | U^{i-1}) + H(S_1 | U_i U^{i-1}) = H(S_1 | U^{i-1}) + H(U_i | U^{i-1} S_1) .$$

As discussed, an immediate consequence of Lemma 11 is that $H(U_i | U^{i-1} S_1) = 1/2$. Thus,

$$|H(U_i | U^{i-1}) - 1/2| = |H(S_1 | U^{i-1}) - H(S_1 | U_i U^{i-1})| .$$

By Lemma 12, there exists an $\epsilon_N \rightarrow 0$ such that

$$0 \leq H(S_1 | U_i U^{i-1}) \leq H(S_1 | U^{i-1}) \leq \epsilon_N .$$

Hence,

$$|H(U_i | U^{i-1}) - 1/2| \leq \epsilon_N .$$

■

VII. APPENDIX

Proof of Corollary 8: By marginalizing (8) over v_i we deduce that

$$p_{\tilde{U}_i|Q_i R_i}(u_i|q_i r_i) = p_{\tilde{U}_i|Q_i}(u_i|q_i). \quad (20)$$

Similarly,

$$p_{\tilde{V}_i|Q_i R_i}(v_i|q_i r_i) = p_{\tilde{V}_i|R_i}(v_i|r_i). \quad (21)$$

Thus, by (8) and the above we deduce that \tilde{U}_i and \tilde{V}_i are independent given Q_i and R_i ,

$$p_{\tilde{U}_i, \tilde{V}_i|Q_i R_i}(u_i v_i|q_i r_i) = p_{\tilde{U}_i|Q_i R_i}(u_i|q_i r_i) \cdot p_{\tilde{V}_i|Q_i R_i}(v_i|q_i r_i). \quad (22)$$

Define

$$h_2(\alpha) = -\alpha \log_2 \alpha - (1 - \alpha) \log_2 (1 - \alpha). \quad (23)$$

We start with the following simple claim: for α, β between 0 and 1,

$$|h_2(\beta) - h_2(\alpha)| \leq h_2(|\beta - \alpha|). \quad (24)$$

Indeed, assume w.l.o.g. that $\beta \geq \alpha$. Then,

$$h_2(\beta) - h_2(\alpha) = \int_{\alpha}^{\beta} h_2'(t) dt \leq \int_0^{\beta - \alpha} h_2'(t) dt = h_2(\beta - \alpha), \quad (25)$$

where the inequality follows from the concavity of h_2 (the derivative h_2' is decreasing). Similarly,

$$h_2(\beta) - h_2(\alpha) = \int_{\alpha}^{\beta} h_2'(t) dt \geq \int_{1 - (\beta - \alpha)}^1 h_2'(t) dt = -h_2(1 - (\beta - \alpha)) = -h_2(\beta - \alpha). \quad (26)$$

We deduce (24) from (25) and (26).

For q_i and r_i fixed, let us adopt the shorthand $\alpha = p_{U_i + V_i|Q_i R_i}(0|q_i r_i)$ and $\beta = p_{\tilde{U}_i + \tilde{V}_i|Q_i R_i}(0|q_i r_i)$. We claim that

$$\begin{aligned} |H(\tilde{U}_i + \tilde{V}_i|Q_i R_i) - H(U_i + V_i|Q_i R_i)| &= \left| \sum_{q_i, r_i} p_{Q_i R_i}(q_i, r_i) (h_2(\beta) - h_2(\alpha)) \right| \\ &\leq \sum_{q_i, r_i} p_{Q_i R_i}(q_i, r_i) |h_2(\beta) - h_2(\alpha)| \\ &\leq \sum_{q_i, r_i} p_{Q_i R_i}(q_i, r_i) h_2(|\beta - \alpha|) \\ &\leq h_2 \left(\sum_{q_i, r_i} p_{Q_i R_i}(q_i, r_i) |\beta - \alpha| \right). \end{aligned} \quad (27)$$

The second inequality follows from (24) while the third inequality follows by applying Jensen's inequality [6, Theorem 2.6.2] with respect to the concave function h_2 .

Our aim now is to bound the argument of h_2 in the RHS of the above displayed equation. Let us use the shorthand $p = p_{U_i V_i|Q_i R_i}$ and $\tilde{p} = p_{\tilde{U}_i \tilde{V}_i|Q_i R_i}$. By (22),

$$I(U_i; V_i|Q_i, R_i) = \sum_{q_i, r_i} p_{Q_i R_i}(q_i, r_i) D(p||\tilde{p}),$$

where $D(p||\tilde{p})$ is the relative entropy between p and \tilde{p} , for q_i and r_i fixed,

$$D(p||\tilde{p}) = \sum_{u_i, v_i} p(u_i, v_i|q_i, r_i) \log_2 \frac{p(u_i, v_i|q_i, r_i)}{\tilde{p}(u_i, v_i|q_i, r_i)}.$$

Next, let us denote $p_+ = p_{U_i + V_i|Q_i R_i}$ and $\tilde{p}_+ = p_{\tilde{U}_i + \tilde{V}_i|Q_i R_i}$. Obviously, p_+ is gotten by quantizing p :

$$p_+(0|q_i r_i) = p(0, 0|q_i r_i) + p(1, 1|q_i r_i), \quad p_+(1|q_i r_i) = p(1, 0|q_i r_i) + p(0, 1|q_i r_i).$$

The same quantization is used to derive \tilde{p}_+ from \tilde{p} . A simple consequence of the log-sum inequality [6, Theorem 2.7.1] is that such a quantization reduces the relative entropy. Namely, for q_i, r_i fixed,

$$D(p||\tilde{p}) \geq D(p_+||\tilde{p}_+).$$

Recalling that $\alpha = p_+(0|q_i r_i)$ and $\beta = \tilde{p}_+(0|q_i r_i)$, we get from Pinsker's inequality [6, Equation 11.147] that

$$D(p_+||\tilde{p}_+) \geq \frac{1}{2 \ln 2} \cdot 2(\beta - \alpha)^2.$$

Aggregating the above inequalities yields

$$I(U_i; V_i | Q_i, R_i) \geq \frac{1}{\ln 2} \sum_{q_i, r_i} p_{Q_i R_i}(q_i, r_i) \cdot (\beta - \alpha)^2 .$$

Now is the time to invoke Lemma 6. Namely, for an ϵ' which we will determine shortly, the fraction of indices i for which $I(U_i; V_i | Q_i, R_i) \leq \epsilon'$ approaches 1 as $N \rightarrow \infty$. Thus, for such an index i we have that

$$\frac{1}{\ln 2} \sum_{q_i, r_i} p_{Q_i R_i}(q_i, r_i) \cdot |\beta - \alpha|^2 \leq \epsilon' .$$

Since squaring is a convex function, we apply Jensen's inequality and deduce that

$$\sum_{q_i, r_i} p_{Q_i R_i}(q_i, r_i) \cdot |\beta - \alpha| \leq \sqrt{\epsilon' \cdot \ln 2} .$$

Assuming the RHS of the above is less than $1/2$, we deduce from the above and from (27) that

$$|H(\tilde{U}_i + \tilde{V}_i | Q_i R_i) - H(U_i + V_i | Q_i R_i)| \leq h_2(\sqrt{\epsilon' \cdot \ln 2}) .$$

Thus, taking ϵ' small enough so that $\sqrt{\epsilon' \cdot \ln 2} \leq 1/2$ and $h_2(\sqrt{\epsilon' \cdot \ln 2}) \leq \epsilon$ finishes the proof. ■

Proof of Lemma 9: Denote the distributions of A and B as

$$A \sim \text{Ber}(\alpha) , \quad B \sim \text{Ber}(\beta) .$$

We will assume w.l.o.g. that $0 \leq \alpha \leq \beta \leq 1/2$. Thus, according to our assumptions,

$$h_2(\alpha) \leq h_2(\beta) , \quad h_2(\alpha) \leq 1 - \xi , \quad h_2(\beta) \geq \xi ,$$

where h_2 is defined in (23). Since h_2 is strictly increasing when restricted to the domain $[0, 1/2]$, it is invertible and we conclude that

$$0 \leq \alpha \leq h_2^{-1}(1 - \xi) , \quad h_2^{-1}(\xi) \leq \beta \leq \frac{1}{2} .$$

We simplify the above to

$$0 \leq \alpha \leq \frac{1}{2} - \sigma , \quad \sigma \leq \beta \leq \frac{1}{2} . \tag{28}$$

where

$$\sigma = \sigma(\xi) = \min \left\{ h_2^{-1}(\xi), \frac{1}{2} - h_2^{-1}(1 - \xi) \right\} .$$

Define the random variable $D = (C, T)$ as follows,

$$D = (C, T) , \quad T \sim \text{Ber}(1/2) , \quad C = \begin{cases} A & \text{if } T = 0 , \\ B & \text{if } T = 1 . \end{cases}$$

One easily gets that

$$H(A + B | D) = \frac{H(A) + H(B)}{2} .$$

Thus, we are interested in bounding the difference

$$H(A + B) - H(A + B | D) = I(A + B; D) .$$

We write $I(X + Y; D)$ as in terms of relative entropy [6, Equation (2.29)], and lower bound that with Pinsker's inequality [6, Equation 11.147]. Doing so results in a straightforward calculation which yields

$$\begin{aligned} H(A + B) - \frac{H(A) + H(B)}{2} &\geq \frac{2}{\ln 2} (\beta(1 - \beta)|1 - 2\alpha| + \alpha(1 - \alpha)|1 - 2\beta|)^2 \\ &\geq \frac{2}{\ln 2} (\beta(1 - \beta)|1 - 2\alpha|)^2 \\ &\geq \frac{2}{\ln 2} (\sigma(1 - \sigma)|\sigma|)^2 \\ &= \frac{2}{\ln 2} \sigma^4 (1 - \sigma)^2 , \end{aligned}$$

where the last inequality follows from (28). Now, simply take Δ as the RHS of the above. ■

REFERENCES

- [1] E. Arkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, pp. 3051–3073, 2009.
- [2] R. Wang, J. Honda, H. Yamamoto, R. Liu, and Y. Hou, "Construction of polar codes for channels with memory," in *Proc. IEEE Inform. Theory Workshop (ITW'2015)*, Jeju Island, Korea, 2015, pp. 187–191.
- [3] E. Şaşoğlu, "Polar coding theorems for discrete systems," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, 2011.
- [4] P. C. Shields, *The Ergodic Theory of Discrete Sample Paths*, ser. Graduate Studies in Mathematics. Providence (R.I.): American Mathematical Society, 1996, vol. 13.
- [5] R. G. Gallager, *Information Theory and Reliable Communications*. New York: John Wiley, 1968.
- [6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.
- [7] E. Arkan and E. Telatar, "On the rate of channel polarization," in *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2009)*, Seoul, South Korea, 2009, pp. 1493–1495.
- [8] E. Arkan, "Source polarization," in *Proc. IEEE Int'l Symp. Inform. Theory (ISIT'2010)*, Austin, Texas, 2010, pp. 899–903.